Statistics and Applications {ISSN 2454-7395 (online)}
Volume 23, No. 2, 2025 (New Series), pp 37–57
http://www.ssca.org.in/journal

Finite Mixtures of Matrix Variate Log-normal Distributions for Clustering Skewed Three-Way Data

Shiva Kumar Kurva and Kiruthika C.
Department of Statistics, Pondicherry University, India

Received: 24 April 2024; Revised: 12 August 2024; Accepted: 03 September 2024

Abstract

The increasing complexity in the dimensionality of data throws new challenges in modelling. Matrix variate data with underlying subpopulations need more sophisticated models to make meaningful statistical inferences and clustering results. This work introduces the finite mixtures of matrix variate log-normal distributions to model the right skewed multimodal matrix variate data, and its application to model-based clustering is discussed. In addition, an extended K-means algorithm is developed as an alternative to the ordinary K-means approach for clustering matrix variate data. Furthermore, it can also be a useful initialization approach for matrix variate finite-mixture models, which received much attention lately for modelling this kind of data. Using the suggested initialization approach, the Expectation-Maximization algorithm is employed to estimate the parameters. The ability of the proposed methodology is illustrated through simulations and real data studies.

Key words: Finite-mixtures; Matrix variate; Log-normal; Model-based clustering; Expectation Maximization; Extended K-means.

AMS Subject Classifications: 91G20, 62G07

1. Introduction

As the dimensionality of data increases, traditional modelling approaches often struggle to capture the intricacies and variations present in complex data sets. One such challenge is modelling three-way or matrix variate data (Viroli, 2011b) where each observation is represented as a matrix rather than a scalar or vector. The multivariate longitudinal data is an example of three-way data. On the other hand, to model and cluster the multi-modal data, the most used characterization is the finite mixture models which are a statistical approach that combines multiple probability distributions to create a more flexible and versatile model. These models are particularly useful when dealing with datasets containing distinct subpopulations or clusters with different statistical characteristics. Vast literature can be seen in finite mixture models. For example, see Everitt (2013), Peel and MacLahlan (2000), McNicholas and Murphy (2008), Bouguila and ElGuebaly (2009), Melnykov and Maitra (2010), McLachlan *et al.* (2019).

Corresponding Author: Shiva Kumar Kurva
Email: shivakumar.sk137@pondiuni.ac.in

In the recent past, high-dimensional multi-modal data was modelled using finite mixtures of matrix variate distributions. Some notable works are by (Viroli, 2011a), Doğru *et al.* (2016), Gallaugher and McNicholas (2018), Tomarchio *et al.* (2022), Tomarchio *et al.* (2020), and Silva *et al.* (2023). The matrix variate finite mixture models found in the literature mostly rely on symmetricity and heavy-tailed distributions. However, the observations at the tails of the distribution may be regarded as outliers in the event of highly skewed data, making it challenging to cluster these observations. Therefore, mixture models with asymmetric distributions are required. Nevertheless, the data that is right-skewed can be modelled by the log-normal distribution. Furthermore, the extreme observations—which are typically considered outliers—are also explained by this distribution. The univariate and multivariate finite mixtures of log-normal distributions are introduced by Deepana and Kiruthika (2018, 2022). Moving forward, we employ matrix variate log-normal distributions as a finite mixtures to model and cluster the matrix variate right-skewed data by utilizing the properties of log-normal distribution, such as skewness and heavy tails.

The parameter estimation of finite mixtures involves identifying latent variable problem. The widely used family of Expectation-Maximization (EM) algorithms efficiently estimate parameters in finite mixtures (Dempster *et al.*, 1977). However, the EM algorithm does not guarantee the global maximum and is sensitive to initial values. Many initialization techniques are discussed in the literature. For example, see Biernacki *et al.* (2003), Karlis and Xekalaki (2003), Michael and Melnykov (2016), Panić *et al.* (2020), and Hu (2015). However, all these techniques are restricted to univariate and multivariate finite mixture models. In recent works of finite mixtures of matrix variate distributions, a random multi-start initializations are employed in parameter estimation. On the other hand, to cluster the matrix variate data, the conventional K-means developed for multivariate data cannot be useful.

To overcome these challenges, the finite mixtures of matrix variate log-normal distributions is introduced to model and cluster the skewed matrix variate data. In addition, an extended K-means algorithm is proposed to cluster the matrix variate data and to obtain the initial values of the parameters in finite mixtures of matrix variate distributions. The rest of the paper is organized as follows: Section 2 introduces the extended K-means clustering for matrix variate data. Section 3 consists of finite mixtures of matrix variate log-normal distributions, its maximum likelihood estimation, and the EM algorithm. Section 4 provides detailed simulation studies and section 5 presents a real data study to validate the proposed methodology. The paper concludes with section 6 by summarising the proposed methodology, simulation and real data studies.

2. Extended K-means clustering

In the realm of data analysis, clustering methods are indispensable tools for uncovering hidden patterns within datasets. Among these techniques, K-means clustering stands as a cornerstone, renowned for its simplicity, efficiency, and effectiveness in partitioning data into distinct and non-overlapping clusters. The essence of K-means lies in its ability to iteratively assign data points to clusters based on the similarity of their features while striving to minimize the within-cluster sum of squares. Moreover, its computational efficiency renders it suitable for large-scale data analysis, making it a preferred choice for real-world applications.

Consider a sample of n matrix observations $X_1, X_2, \ldots, X_n$ that are to be segregated into m ($\leq n$) clusters $C_1, C_2, \ldots, C_m$. Let c_j ($j = 1, 2, \ldots, m$) be the centre matrix of the cluster C_j. The essential part of the K-means is to calculate the distance between cluster centres and observations. Rezaei and Ahmadi (2023) have given the distance between two matrix observations X and Y which are observed from a matrix variate distribution is,

$$D(X, Y) = \sqrt{tr\left\{\Psi^{-1}(X - Y)^T\Sigma^{-1}(X - Y)\right\}}. \tag{1}$$

where Ψ and Σ are the positive definite matrices. Here, $tr(.)$ is the trace of a symmetric matrix and $(\cdot)^T$ denotes the transpose of matrix. Considering the above distance by keeping $\Psi = \Sigma = I$, the distance between a matrix observation X_i and a cluster centre matrix c_j is given as,

$$D(X_i, c_j) = \sqrt{tr\left\{(X_i - c_j)^T(X_i - c_j)\right\}}. \tag{2}$$

Once the initial centres are identified and the distance between each observation and the centres is obtained, the observations are to be segregated into updated cluster centres. For that, the updated cluster centres are which minimize the within-cluster sum of squares, which is,

$$W = \sum_{j=1}^{m} \sum_{X_i \in C_j} tr\left\{(X_i - c_j)^T(X_i - c_j)\right\}. \tag{3}$$

On minimizing the within-cluster sum of squares, the least-squares estimates of the updated cluster centres are obtained as,

$$c_j = \frac{1}{n_j} \sum_{X_i \in C_j} X_i \tag{4}$$

where n_j is the size of cluster C_j.

Since the results of the K-means depends on initial centres, the algorithm needs multiple compilations with random initial centres, typically observations of the data. The best clustering solution is the one that has the lowest within-cluster sum of squares. This method provides the clustered observations, corresponding cluster centres, and sizes. The step-by-step procedure is given in Algorithm 1. This clustering technique can be used for any matrix variate data which is reasonably well separated. Furthermore, the location parameters and mixing proportions of the finite mixtures of matrix variate distributions can be initialized with these cluster centres and cluster sizes, respectively. In further studies, we focus on alternative distance metrics and different initializations for the extended K-means technique. In the next section, the proposed technique is used as an initialization to the finite mixtures of matrix variate log-normal distributions.

3. Finite mixtures of matrix variate Log-normal distributions

For many practical reasons, it is obvious to observe data on multiple variables (say columns) at different time points or situations (say rows). As a result, there are two covariance structures, one along columns and the other along rows. Consider a random matrix $\mathcal{Y}$ of

Algorithm 1 Extended K-means algorithm

procedure K-means(X, m)
  for $X = [x_i]; i = 1, \ldots, n$ & $C_j; j = 1, \ldots, m$ **do** ▷ Initialization
    Random Initialize of c_j (j^{th} Cluster Centre) from X
  end for
  while $c_j^{(t)} \neq c_j^{(t+1)}$ **do** ▷ Core
    for $X = [x_i]; i = 1, \ldots, n$ & $C_j; j = 1, \ldots, m$ **do**
      update $I\left(x_i \in C_j\right) \forall\ j = \arg\min_j d(X_i, c_j)$ using (2)
      update c_j according to (4).
    end for
  end while
  return $c_j, n_j, I\left(X_i \in C_j\right)$
end procedure

order $(r \times c)$, assumed to be distributed as matrix variate Log-normal (MVLN) distribution (Rohde *et al.*, 2012) with the mean matrix M of order $(r \times c)$, the row covariance matrix Ω of order $(r \times r)$, and the column covariance matrix Σ of order $(c \times c)$. The probability density function of $\mathcal{Y}$ is given by,

$$f(Y; M, \Omega, \Sigma) = \frac{1}{(2\pi)^{\frac{rc}{2}} |\Sigma|^{\frac{r}{2}} |\Omega|^{\frac{c}{2}} \prod_{p=1}^{r} \prod_{q=1}^{c} y_{pq}} \times \exp\left\{-\frac{1}{2}\mathrm{tr}\left[\Sigma^{-1}(\log Y - M)\Omega^{-1}(\log Y - M)^T\right]\right\} \quad (5)$$

where y_{pq} is the value of p^{th} row and q^{th} column of an observation matrix Y. Here, Ω and Σ are positive definite matrices. One of the advantages of this distribution is that one can obtain the closed-form expressions for the parameter estimates.

Let $\mathcal{X}$ be a random matrix that has finite mixtures of matrix variate log-normal (FMMVLN) distributions, the pdf of $\mathcal{X}$ is given as,

$$\begin{aligned} g(X; \Theta) &= \sum_{j=1}^{k} \pi_j f(X; M_j, \Omega_j, \Sigma_j) \\ &= \sum_{j=1}^{k} \pi_j \frac{1}{(2\pi)^{\frac{rc}{2}} |\Sigma_j|^{\frac{r}{2}} |\Omega_j|^{\frac{c}{2}} \prod_{p=1}^{r} \prod_{q=1}^{c} x_{pq}} \times \\ &\qquad \exp\left\{-\frac{1}{2}\mathrm{tr}\left[\Sigma_j^{-1}(\log X - M_j)\Omega_j^{-1}(\log X - M_j)^T\right]\right\} \end{aligned} \quad (6)$$

where $\pi_j (0 < \pi_j \leq 1)$ is the mixing proportion of the j^{th} component density $f(X; M_j, \Omega_j, \Sigma_j)$. The prior probability that an observed matrix belongs to each component density is π_j $(j = 1, 2, \ldots, k)$ and $\sum_{j=1}^{k} \pi_j = 1$. Here, $\Theta = (\pi_j, M_j, \Omega_j, \Sigma_j; j = 1, 2, \ldots, k)$ is the parametric space of the mixture density.

3.1. Maximum likelihood estimation

Consider a sample $X_1, X_2, \ldots, X_n$ (say $\mathbf{X}$) of size n assumed to be drawn from (6), the log-likelihood function is,

$$l(\Theta|\mathbf{X}) = \sum_{i=1}^{n} \log \left\{ \sum_{j=1}^{k} \pi_j f(X_i; M_j, \Omega_j, \Sigma_j) \right\}. \tag{7}$$

Since the summation is involved within the *'log'* function, estimation of parameters through (7) is difficult. The EM algorithm is employed to overcome this and estimate parameters efficiently. Here, the observed data is incomplete as there is no information about the component membership of observations. Let z be a multinomial random variable that denotes from which component density each observation came. *i.e.*,

$$z_{ij} = \begin{cases} 1 & ; \text{ if } X_i \text{ came from } j^{th} \text{ component density.} \\ 0 & ; \text{ elsewhere.} \end{cases}$$

Here, $\sum_{j=1}^{k} z_{ij} = 1$. From the definition of finite mixture density,

$$f(z_{ij} = 1) = \pi_j \quad \text{and} \quad g(X_i|z_{ij} = 1) = f(X_i; M_j, \Omega_j, \Sigma_j). \tag{8}$$

Further, using the marginal distribution of z_{ij} and the conditional distribution of $X_i|z_{ij} = 1$, the joint density of the complete data (X_i, z_i) can be written as,

$$\begin{aligned} f(X, z) &= f(z_{ij} = 1) g(X_i|z_{ij} = 1) \\ &= \prod_{j=1}^{k} \{\pi_j f(X_i; M_j, \Omega_j, \Sigma_j)\}^{z_{ij}} \end{aligned} \tag{9}$$

By utilizing the information about z_{ij}, the summation in the mixture density replaced by the product, which makes the further parameter estimation easy. Using (9), for a random sample of size n, the log-likelihood function of the complete-data becomes,

$$\begin{aligned} l(\Theta|\mathbf{X}, z) &= \sum_{i=1}^{n} \sum_{j=1}^{k} z_{ij} \log \{\pi_j f(X_i; M_j, \Omega_j, \Sigma_j)\} \\ &= \sum_{i=1}^{n} \sum_{j=1}^{k} z_{ij} \log \Bigg\{ \pi_j \frac{1}{(2\pi)^{\frac{rc}{2}} |\Sigma_j|^{\frac{r}{2}} |\Omega_j|^{\frac{c}{2}} \prod_{p=1}^{r} \prod_{q=1}^{c} x_{pq}} \times \\ &\qquad \exp \left\{ -\frac{1}{2} \mathrm{tr} \left[\Sigma_j^{-1} (\log X_i - M_j) \Omega_j^{-1} (\log X_i - M_j)^T \right] \right\} \Bigg\} \end{aligned} \tag{10}$$

3.2. EM algorithm

The maximization of (10) can be done through EM-algorithm given in Algorithm 2. Each iteration of E-step calculates $Q(\Theta, \Theta^{(t)})$ which is the expected value of the log-likelihood

Algorithm 2 EM algorithm: FMMVLN distributions

procedure FMMVLN(X, k)
  for $X = \{X_i\}_{i=1}^{n}$ & $j = 1, \ldots, k$ **do** ▷ Initialization
    Initialize M_j & π_j from Extended K-means.
    Set Ω_j & Σ_j as identity matrices.
  end for
  while $M_j^{(t+1)} \neq M_j^{(t)}$ **do**
    for $j = 1, \ldots, k$ **do**
      Update γ_{ij} according to (11). ▷ E-step
      Update $(\widehat{M_j},\ \widehat{\Omega_j},\ \widehat{\Sigma_j},\ \widehat{\pi_j})$ according to (12). ▷ M-step
    end for
  end while
  Calculate AIC and BIC.
return $M_j,\ \Omega_j,\ \Sigma_j,\ \pi_j, \gamma_{ij}$, AIC, and BIC.
end procedure

of complete data w.r.t the conditional distribution of latent variable z, given the observed data, at the current value of Θ as $\Theta^{(t)}$.

Using Bayes theorem, the posterior probability that an observed matrix belongs to the j^{th} component density can be written as,

$$\begin{aligned} E\left[z_{ij}|\mathcal{X};\Theta^{(t)}\right] &= f\left(z_{ij}=1|X_i;\Theta^{(t)}\right) \\ &= \frac{f\left(z_{ij}=1\right) g\left(X_i;\Theta^{(t)}|z_{ij}=1\right)}{g\left(X_i;\Theta^{(t)}\right)} \\ &= \frac{\pi_j^{(t)} f\left(X_i;M_j^{(t)},\Omega_j^{(t)},\Sigma_j^{(t)}\right)}{\sum_{j=1}^{k}\pi_j^{(t)} f\left(X_i;M_j^{(t)},\Omega_j^{(t)},\Sigma_j^{(t)}\right)} = \gamma_{ij}^{(t)} \quad \text{(say)}. \end{aligned} \tag{11}$$

Now, the expected value of the complete data log-likelihood, $Q(\Theta, \Theta^{(t)})$ is given as,

$$\begin{aligned} Q(\Theta,\Theta^{(t)}) &= E_{z|X_i;\Theta^{(t)}}\left\{\sum_{i=1}^{n}\sum_{j=1}^{k} z_{ij}\log\left\{\pi_j f(X_i;M_j,\Omega_j,\Sigma_j)\right\}\right\} \\ &= \sum_{i=1}^{n}\sum_{j=1}^{k} E[z_{ij}|X_i;\Theta^{(t)}]\log\left\{\pi_j f(X_i;M_j,\Omega_j,\Sigma_j\right\} \\ &= \sum_{i=1}^{n}\sum_{j=1}^{k}\gamma_{ij}^{(t)}\Big\{\log\pi_j - \frac{rc}{2}\log(2\pi) - \frac{c}{2}\log|\Omega_j| - \frac{r}{2}\log|\Sigma_j| \\ &\quad - \sum_{p=1}^{r}\sum_{q=1}^{c} x_{pq} - \frac{1}{2}\mathrm{tr}\left[\Sigma_j^{-1}(\log X_i - M_j)\Omega_j^{-1}(\log X_i - M_j)^T\right]\Big\} \end{aligned}$$

The M-step maximizes $Q(\Theta, \Theta^{(t)})$ by obtaining derivatives with respect to the mixture parameters and equating to zero. Here, the estimates of π_j are obtained under the constraints

$0 < \pi_j < 1$ and $\sum_{j=1}^{k} \pi_j = 1$. The maximum likelihood estimates are given below:

$$\begin{aligned}
\widehat{M}_j^{(t+1)} &= \frac{\sum_{i=1}^{n} \gamma_{ij}^{(t)} \log X_i}{\sum_{i=1}^{n} \gamma_{ij}^{(t)}} \\
\widehat{\Omega}_j^{(t+1)} &= \frac{\sum_{i=1}^{n} \gamma_{ij}^{(t)} (\log X_i - \widehat{M}_j^{(t+1)})\ \widehat{\Sigma}_j^{(t)^{-1}} (\log X_i - \widehat{M}_j^{(t+1)})^T}{c \sum_{i=1}^{n} \gamma_{ij}^{(t)}} \\
\widehat{\Sigma}_j^{(t+1)} &= \frac{\sum_{i=1}^{n} \gamma_{ij}^{(t)} (\log X_i - \widehat{M}_j^{(t+1)})^T\ \widehat{\Omega}_j^{(t+1)^{-1}} (\log X_i - \widehat{M}_j^{(t+1)})}{r \sum_{i=1}^{n} \gamma_{ij}^{(t)}} \\
\widehat{\pi}_j^{(t+1)} &= \frac{\sum_{i=1}^{n} \gamma_{ij}^{(t)}}{n}
\end{aligned} \tag{12}$$

The E and M steps are iteratively repeated until the convergence is reached. The step-by-step procedure of EM algorithm is given in Algorithm 2. In the context of finite mixture models, initialization is the major challenge, as the estimates of the EM algorithm are sensitive to the initialization of parameter values. Different starting values can provide different local maxima and convergence rates of the EM algorithm. Consequently, the clustering results also vary; hence, having a unique clustering solution is difficult. The next section deals with the necessary illustrations of the proposed methodology with simulated and real data studies.

4. Simulation studies

4.1. Extended K-means clustering

To evaluate the performance of the extended K-means clustering, we considered three different sample sizes: 250, 500, and 750, with two and three clusters. The data dimension for two clusters is considered as 2×2 and 3×4 for three clusters. A total of 100 datasets were generated from matrix variate normal populations, with mean matrices chosen arbitrarily to maintain some degree of overlapping. The covariance matrices were set as identity matrices. We then varied the proportion of cluster sizes to create different scenarios for each combination. The extended K-means algorithm was applied to all 100 datasets ranging the cluster size from 1 to 6, and the within sum of squares are evaluated. The average within sum of squares of 100 datasets are plotted against the number of clusters.

The optimal number of clusters are chosen based on the 'elbow' of scree plot which indicates that adding more clusters does not significantly decrease the within sum of squares. The clustering performance is evaluated using Adjusted Rand Index (ARI) (Rand, 1971), a widely used metric which provides a comprehensive assessment of the clustering quality. The ARI value ranges from 0 to 1, where 0 indicates the poorest classification and 1 represents perfect classification. From Figure 1, the scree plots in the first row shows that the optimal clusters size is two while the second row reveals that there are three clusters. Hence, the results with corresponding cluster sizes are given Table 1. From these results, the estimated proportion of cluster sizes and centre matrices closely match the real parameter values, indicating that the extended K-means is adequate for parameter estimation. Besides, for a fixed overlapping between clusters, the ARI values of 2×2 data are increasing with sam-

ple size. However, ARI values of 3×4 data are substantially closer to 1 across all sample sizes, indicating that the performance of clustering improves as the sample size and data dimension increases. These outcomes demonstrate the suggested extended K-means method's effectiveness in clustering. Further, as the sample size grows, the standard deviation of every estimate drops, showing the consistency of the estimates.

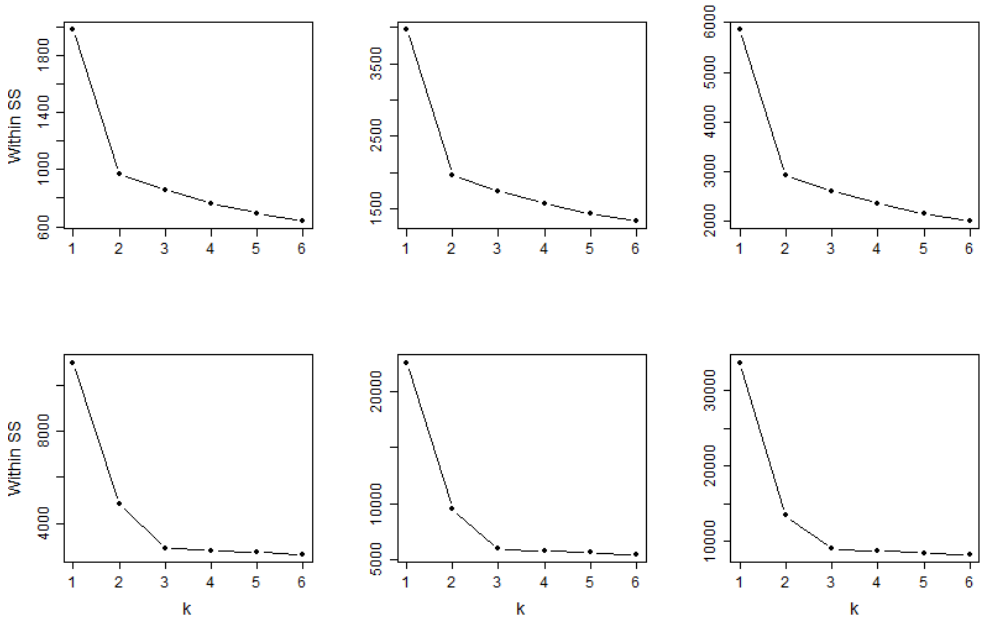

Figure 1: Scree plots for within-cluster sum of squares against number of clusters. The first row represents two cluster data and the second row represents three cluster data

4.2. Finite mixtures of MVLN distributions

In order to exhibit the behaviour of the MVLN mixture model, illustrations are given with two and three-component mixture models and component densities are generated. All the simulations under the two-component mixture model are done at 250, 500, and 750 sample sizes, whereas under the three-component mixture model, simulations are performed with 500 and 750 sample sizes. Here, 100 samples generated from each of the parameter combinations.

A two-component FMMVLN model with data dimension of (2×3) is considered and different combination of parameters are used to generate the samples for each sample size given in Table 2. The mean matrices of all models are chosen arbitrarily to maintain some degree of overlapping. For the sample size of 250, the covariance matrices within clusters considered as homogenous. For the sample size of 500, the row and column covariance matrices between clusters are kept homogenous. For the sample size of 750, the row and column covariance matrices within and between clusters considered heterogeneous. For the case of three-component FMMVLN model, two sample sizes 500 and 750 with the data dimension of (3×3) is considered. A similar approach to the two-component model is considered to generate mean and covariance matrices.

Table 1: Results of extended K-means clustering for simulated data

k	n	C	λ	c	$\widehat{\lambda}$ (sd)	$\widehat{c}$ (sd)	ARI (sd)
2	250	1	0.5	$\mathbf{0}_{22}$	0.499 (0.013)	$\begin{bmatrix}-0.011 & 0.004\\ -0.007 & -0.007\end{bmatrix}\left(\begin{bmatrix}0.093 & 0.103\\ 0.092 & 0.102\end{bmatrix}\right)$	0.904 (0.038)
		2	0.5	$2\ \mathbf{J}_{22}$	0.501 (0.013)	$\begin{bmatrix}2.002 & 2.004\\ 1.992 & 2.007\end{bmatrix}\left(\begin{bmatrix}0.095 & 0.094\\ 0.093 & 0.102\end{bmatrix}\right)$	
	500	1	0.45	$\mathbf{0}_{22}$	0.453 (0.007)	$\begin{bmatrix}-0.004 & 0.005\\ -0.004 & 0.001\end{bmatrix}\left(\begin{bmatrix}0.073 & 0.078\\ 0.073 & 0.074\end{bmatrix}\right)$	0.911 (0.026)
		2	0.55	$2\ \mathbf{J}_{22}$	0.547 (0.007)	$\begin{bmatrix}2.020 & 2.021\\ 2.002 & 2.015\end{bmatrix}\left(\begin{bmatrix}0.063 & 0.067\\ 0.064 & 0.061\end{bmatrix}\right)$	
	750	1	0.4	$\mathbf{0}_{22}$	0.406 (0.006)	$\begin{bmatrix}0.006 & 0.008\\ 0.002 & 0.007\end{bmatrix}\left(\begin{bmatrix}0.061 & 0.069\\ 0.063 & 0.067\end{bmatrix}\right)$	0.911 (0.020)
		2	0.6	$2\ \mathbf{J}_{22}$	0.594 (0.006)	$\begin{bmatrix}2.022 & 2.014\\ 2.010 & 2.020\end{bmatrix}\left(\begin{bmatrix}0.052 & 0.050\\ 0.048 & 0.051\end{bmatrix}\right)$	
3	250	1	0.33	$\mathbf{0}_{34}$	0.329 (0.001)	$\begin{bmatrix}0.001 & -0.009 & -0.008 & -0.013\\ 0.004 & -0.006 & 0.011 & 0.007\\ 0.022 & -0.013 & -0.007 & 0.005\end{bmatrix}\left(\begin{bmatrix}0.112 & 0.108 & 0.116 & 0.111\\ 0.109 & 0.112 & 0.103 & 0.121\\ 0.118 & 0.122 & 0.109 & 0.110\end{bmatrix}\right)$	1 (0.002)
		2	0.33	$2\ \mathbf{J}_{34}$	0.330 (0.001)	$\begin{bmatrix}1.982 & 1.984 & 2.009 & 1.994\\ 1.982 & 2.029 & 1.998 & 2.008\\ 2.009 & 2.000 & 2.009 & 2.017\end{bmatrix}\left(\begin{bmatrix}0.112 & 0.123 & 0.116 & 0.110\\ 0.104 & 0.110 & 0.108 & 0.107\\ 0.111 & 0.117 & 0.123 & 0.094\end{bmatrix}\right)$	
		3	0.34	$4\ \mathbf{J}_{34}$	0.341 (0.001)	$\begin{bmatrix}3.992 & 4.001 & 3.974 & 4.012\\ 3.985 & 4.014 & 4.002 & 3.988\\ 3.984 & 3.993 & 4.006 & 4.007\end{bmatrix}\left(\begin{bmatrix}0.115 & 0.110 & 0.102 & 0.098\\ 0.126 & 0.100 & 0.104 & 0.097\\ 0.107 & 0.109 & 0.109 & 0.104\end{bmatrix}\right)$	
	500	1	0.4	$\mathbf{0}_{34}$	0.4 (0.000)	$\begin{bmatrix}-0.011 & -0.007 & -0.001 & -0.004\\ -0.006 & 0.015 & 0.003 & 0.002\\ 0.006 & -0.011 & 0.002 & 0.012\end{bmatrix}\left(\begin{bmatrix}0.070 & 0.075 & 0.073 & 0.071\\ 0.068 & 0.069 & 0.066 & 0.068\\ 0.079 & 0.077 & 0.073 & 0.070\end{bmatrix}\right)$	0.999 (0.002)
		2	0.3	$2\ \mathbf{J}_{34}$	0.3 (0.001)	$\begin{bmatrix}2.007 & 2.009 & 1.993 & 2.010\\ 2.001 & 1.997 & 1.994 & 1.996\\ 2.002 & 2.000 & 2.008 & 2.004\end{bmatrix}\left(\begin{bmatrix}0.081 & 0.087 & 0.074 & 0.079\\ 0.072 & 0.085 & 0.083 & 0.076\\ 0.075 & 0.071 & 0.082 & 0.076\end{bmatrix}\right)$	
		3	0.3	$4\ \mathbf{J}_{34}$	0.3 (0.001)	$\begin{bmatrix}4.003 & 3.995 & 3.998 & 4.011\\ 4.014 & 4.007 & 3.997 & 3.985\\ 4.006 & 4.005 & 3.996 & 3.997\end{bmatrix}\left(\begin{bmatrix}0.077 & 0.076 & 0.084 & 0.083\\ 0.086 & 0.078 & 0.079 & 0.076\\ 0.086 & 0.077 & 0.084 & 0.085\end{bmatrix}\right)$	
	750	1	0.50	$\mathbf{0}_{34}$	0.499 (0.001)	$\begin{bmatrix}-0.002 & -0.001 & -0.004 & 0.001\\ -0.001 & 0.005 & 0.000 & -0.003\\ 0.003 & -0.006 & 0.004 & 0.008\end{bmatrix}\left(\begin{bmatrix}0.052 & 0.059 & 0.046 & 0.055\\ 0.050 & 0.056 & 0.051 & 0.045\\ 0.054 & 0.053 & 0.054 & 0.048\end{bmatrix}\right)$	0.999 (0.002)
		2	0.25	$2\ \mathbf{J}_{34}$	0.251 (0.001)	$\begin{bmatrix}1.996 & 1.996 & 1.994 & 2.018\\ 2.009 & 2.010 & 1.996 & 1.995\\ 2.001 & 2.003 & 1.994 & 1.995\end{bmatrix}\left(\begin{bmatrix}0.072 & 0.067 & 0.069 & 0.075\\ 0.075 & 0.077 & 0.073 & 0.072\\ 0.076 & 0.074 & 0.077 & 0.074\end{bmatrix}\right)$	
		3	0.25	$4\ \mathbf{J}_{34}$	0.250 (0.000)	$\begin{bmatrix}3.988 & 3.996 & 3.994 & 4.005\\ 4.013 & 3.987 & 3.999 & 3.997\\ 3.989 & 3.996 & 4.000 & 3.993\end{bmatrix}\left(\begin{bmatrix}0.068 & 0.070 & 0.072 & 0.076\\ 0.071 & 0.074 & 0.070 & 0.071\\ 0.070 & 0.068 & 0.079 & 0.072\end{bmatrix}\right)$	

Here, $\mathbf{0}_{ij}$ and $\mathbf{J}_{ij}$ are the null matrix and matrix of ones of order $(i \times j)$, respectively.

To assess the model fit and identify the number of components in model, we use Akaike Information Criterion (AIC) (Akaike, 1974) and Bayesian Information Criterion (BIC) (Schwarz, 1978), the effective and widely used metrics. We consider AIC and BIC formulations where higher values indicate better model fit. From (12), the estimates of location parameters of the model are the functions of '$\log X$'. Hence, the proposed K-means is fitted to '$\log X$', and the initial values are obtained for each sample generated. From the results of extended K-means, the cluster centres and the proportion of cluster sizes are used to initialize the mean matrices and mixing proportions, respectively. The initialization for

both row and column covariance matrices is considered identity matrices of the corresponding order. Then, the FMMVLN model is fitted through the EM algorithm, varying number of components from 1 to 6. We calculated the ARI to evaluate the model's clustering performance and misclassification rate (MCR) to measure the proportion of incorrectly classified observations.

Table 2: Parameters used to simulate data from two and three-component FMMVLN models

K	n	k^{a}	π	M	Ω	Σ
2	250	1	0.35	$\begin{bmatrix}0 & 0 & 0\\ 0 & 0 & 0\end{bmatrix}$	$\begin{bmatrix}0.5 & 0\\ 0 & 0.5\end{bmatrix}$	$\begin{bmatrix}0.5 & 0 & 0\\ 0 & 0.5 & 0\\ 0 & 0 & 0.5\end{bmatrix}$
		2	0.65	$\begin{bmatrix}0.6 & 0.6 & 0.6\\ 0.6 & 0.6 & 0.6\end{bmatrix}$	$\begin{bmatrix}0.75 & 0\\ 0 & 0.75\end{bmatrix}$	$\begin{bmatrix}0.75 & 0 & 0\\ 0 & 0.75 & 0\\ 0 & 0 & 0.75\end{bmatrix}$
	500	1	0.4	$\begin{bmatrix}0.5 & 0.5 & 0.5\\ 0.5 & 0.5 & 0.5\end{bmatrix}$	$\begin{bmatrix}0.75 & 0.25\\ 0.25 & 0.75\end{bmatrix}$	$\begin{bmatrix}1 & 0 & 0.5\\ 0 & 1 & 0.25\\ 0.5 & 0.25 & 1\end{bmatrix}$
		2	0.6	$\begin{bmatrix}1.8 & 1.8 & 1.8\\ 1.8 & 1.8 & 1.8\end{bmatrix}$	$\begin{bmatrix}0.75 & 0.25\\ 0.25 & 0.75\end{bmatrix}$	$\begin{bmatrix}1 & 0 & 0.5\\ 0 & 1 & 0.25\\ 0.5 & 0.25 & 1\end{bmatrix}$
	750	1	0.5	$\begin{bmatrix}1 & 1 & 1\\ 1 & 1 & 1\end{bmatrix}$	$\begin{bmatrix}1 & 0.4\\ 0.4 & 0.75\end{bmatrix}$	$\begin{bmatrix}1 & 0.25 & 0\\ 0.25 & 1 & 0.25\\ 0 & 0.25 & 1\end{bmatrix}$
		2	0.5	$\begin{bmatrix}2.5 & 2.5 & 2.5\\ 2.5 & 2.5 & 2.5\end{bmatrix}$	$\begin{bmatrix}1.25 & 0.25\\ 0.25 & 1\end{bmatrix}$	$\begin{bmatrix}0.75 & 0 & 0.25\\ 0 & 1 & 0.5\\ 0.25 & 0.5 & 1.25\end{bmatrix}$
3	500	1	0.33	$\begin{bmatrix}0 & 0 & 0\\ 0 & 0 & 0\\ 0 & 0 & 0\end{bmatrix}$	$\begin{bmatrix}0.5 & 0 & 0\\ 0 & 0.5 & 0\\ 0 & 0 & 0.5\end{bmatrix}$	$\begin{bmatrix}0.75 & 0 & 0\\ 0 & 0.75 & 0\\ 0 & 0 & 0.75\end{bmatrix}$
		2	0.33	$\begin{bmatrix}0.7 & 0.7 & 0.7\\ 0.7 & 0.7 & 0.7\\ 0.7 & 0.7 & 0.7\end{bmatrix}$	$\begin{bmatrix}0.5 & 0 & 0\\ 0 & 0.5 & 0\\ 0 & 0 & 0.5\end{bmatrix}$	$\begin{bmatrix}0.75 & 0 & 0\\ 0 & 0.75 & 0\\ 0 & 0 & 0.75\end{bmatrix}$
		3	0.34	$\begin{bmatrix}1.4 & 1.4 & 1.4\\ 1.4 & 1.4 & 1.4\\ 1.4 & 1.4 & 1.4\end{bmatrix}$	$\begin{bmatrix}0.5 & 0 & 0\\ 0 & 0.5 & 0\\ 0 & 0 & 0.5\end{bmatrix}$	$\begin{bmatrix}0.75 & 0 & 0\\ 0 & 0.75 & 0\\ 0 & 0 & 0.75\end{bmatrix}$
	750	1	0.30	$\begin{bmatrix}0.5 & 0.5 & 0.5\\ 0.5 & 0.5 & 0.5\\ 0.5 & 0.5 & 0.5\end{bmatrix}$	$\begin{bmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & 1\end{bmatrix}$	$\begin{bmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & 1\end{bmatrix}$
		2	0.30	$\begin{bmatrix}1.2 & 1.2 & 1.2\\ 1.2 & 1.2 & 1.2\\ 1.2 & 1.2 & 1.2\end{bmatrix}$	$\begin{bmatrix}0.75 & 0 & 0\\ 0 & 0.75 & 0\\ 0 & 0 & 0.75\end{bmatrix}$	$\begin{bmatrix}0.75 & 0 & 0\\ 0 & 0.75 & 0\\ 0 & 0 & 0.75\end{bmatrix}$
		3	0.40	$\begin{bmatrix}2 & 2 & 2\\ 2 & 2 & 2\\ 2 & 2 & 2\end{bmatrix}$	$\begin{bmatrix}0.5 & 0 & 0\\ 0 & 0.5 & 0\\ 0 & 0 & 0.5\end{bmatrix}$	$\begin{bmatrix}0.5 & 0 & 0\\ 0 & 0.5 & 0\\ 0 & 0 & 0.5\end{bmatrix}$

[a] represents the component label in mixture model.

Table 3: Number of samples with best AIC and BIC values out of 100 samples for different combinations of K, n, and g

K	2						3			
n	250		500		750		500		750	
g	AIC	BIC	AIC	BIC	AIC	BIC	AIC	BIC	AIC	BIC
1	0	11	0	5	0	0	0	0	0	0
2	52	89	50	95	49	100	0	0	0	0
3	24	0	18	0	16	0	54	100	55	100
4	12	0	11	0	14	0	19	0	26	0
5	8	0	10	0	12	0	13	0	11	0
6	4	0	11	0	9	0	14	0	8	0

Table 4: Estimation results of two-component FMMVLN model

n	k[a]	$\widehat{\pi}$ (sd)	$\widehat{M}$ (sd)	$\widehat{\Omega}$ (sd)	$\widehat{\Sigma}$ (sd)
250	1	0.349	$\begin{bmatrix}-0.001 & -0.004 & -0.002\\ 0.001 & 0.003 & 0.001\end{bmatrix}$	$\begin{bmatrix}0.238 & -0.001\\ -0.001 & 0.248\end{bmatrix}$	$\begin{bmatrix}1.007 & -0.017 & -0.014\\ -0.017 & 1.000 & -0.001\\ -0.014 & -0.001 & 1.009\end{bmatrix}$
		(0.038)	$\left(\begin{bmatrix}0.064 & 0.073 & 0.065\\ 0.075 & 0.071 & 0.075\end{bmatrix}\right)$	$\left(\begin{bmatrix}0.031 & 0.023\\ 0.023 & 0.030\end{bmatrix}\right)$	$\left(\begin{bmatrix}0.111 & 0.099 & 0.104\\ 0.099 & 0.120 & 0.109\\ 0.104 & 0.109 & 0.111\end{bmatrix}\right)$
	2	0.651	$\begin{bmatrix}0.594 & 0.598 & 0.599\\ 0.599 & 0.598 & 0.604\end{bmatrix}$	$\begin{bmatrix}0.563 & 0.000\\ 0.000 & 0.558\end{bmatrix}$	$\begin{bmatrix}0.999 & -0.009 & 0.011\\ -0.009 & 0.991 & -0.006\\ 0.011 & -0.006 & 1.001\end{bmatrix}$
		(0.038)	$\left(\begin{bmatrix}0.068 & 0.060 & 0.067\\ 0.067 & 0.069 & 0.067\end{bmatrix}\right)$	$\left(\begin{bmatrix}0.041 & 0.031\\ 0.031 & 0.043\end{bmatrix}\right)$	$\left(\begin{bmatrix}0.077 & 0.061 & 0.059\\ 0.061 & 0.069 & 0.071\\ 0.059 & 0.071 & 0.073\end{bmatrix}\right)$
500	1	0.403	$\begin{bmatrix}0.508 & 0.510 & 0.505\\ 0.502 & 0.504 & 0.496\end{bmatrix}$	$\begin{bmatrix}0.735 & 0.247\\ 0.247 & 0.734\end{bmatrix}$	$\begin{bmatrix}1.020 & -0.005 & 0.509\\ -0.005 & 1.024 & 0.248\\ 0.509 & 0.248 & 1.013\end{bmatrix}$
		(0.040)	$\left(\begin{bmatrix}0.095 & 0.086 & 0.089\\ 0.097 & 0.087 & 0.088\end{bmatrix}\right)$	$\left(\begin{bmatrix}0.059 & 0.047\\ 0.047 & 0.052\end{bmatrix}\right)$	$\left(\begin{bmatrix}0.073 & 0.070 & 0.066\\ 0.070 & 0.071 & 0.065\\ 0.066 & 0.065 & 0.060\end{bmatrix}\right)$
	2	0.597	$\begin{bmatrix}1.802 & 1.808 & 1.804\\ 1.806 & 1.814 & 1.809\end{bmatrix}$	$\begin{bmatrix}0.731 & 0.239\\ 0.239 & 0.727\end{bmatrix}$	$\begin{bmatrix}1.021 & 0.006 & 0.502\\ 0.006 & 1.026 & 0.251\\ 0.502 & 0.251 & 1.012\end{bmatrix}$
		(0.040)	$\left(\begin{bmatrix}0.073 & 0.066 & 0.070\\ 0.071 & 0.076 & 0.073\end{bmatrix}\right)$	$\left(\begin{bmatrix}0.043 & 0.030\\ 0.030 & 0.046\end{bmatrix}\right)$	$\left(\begin{bmatrix}0.048 & 0.043 & 0.045\\ 0.043 & 0.055 & 0.043\\ 0.045 & 0.043 & 0.050\end{bmatrix}\right)$
750	1	0.500	$\begin{bmatrix}1.007 & 0.997 & 0.995\\ 1.006 & 0.993 & 0.998\end{bmatrix}$	$\begin{bmatrix}0.992 & 0.395\\ 0.395 & 0.740\end{bmatrix}$	$\begin{bmatrix}1.008 & 0.250 & 0.004\\ 0.250 & 1.003 & 0.254\\ 0.004 & 0.254 & 1.013\end{bmatrix}$
		(0.020)	$\left(\begin{bmatrix}0.067 & 0.061 & 0.059\\ 0.058 & 0.051 & 0.048\end{bmatrix}\right)$	$\left(\begin{bmatrix}0.045 & 0.028\\ 0.028 & 0.035\end{bmatrix}\right)$	$\left(\begin{bmatrix}0.050 & 0.043 & 0.039\\ 0.043 & 0.049 & 0.041\\ 0.039 & 0.041 & 0.044\end{bmatrix}\right)$
	2	0.500	$\begin{bmatrix}2.498 & 2.498 & 2.497\\ 2.496 & 2.491 & 2.500\end{bmatrix}$	$\begin{bmatrix}1.203 & 0.240\\ 0.240 & 0.959\end{bmatrix}$	$\begin{bmatrix}0.779 & 0.000 & 0.262\\ 0.000 & 1.042 & 0.513\\ 0.262 & 0.513 & 1.291\end{bmatrix}$
		(0.020)	$\left(\begin{bmatrix}0.067 & 0.057 & 0.076\\ 0.053 & 0.070 & 0.073\end{bmatrix}\right)$	$\left(\begin{bmatrix}0.052 & 0.041\\ 0.041 & 0.045\end{bmatrix}\right)$	$\left(\begin{bmatrix}0.039 & 0.042 & 0.039\\ 0.042 & 0.043 & 0.044\\ 0.039 & 0.044 & 0.057\end{bmatrix}\right)$

[a] represents the component label in mixture model.

Table 5: Estimation results of three-component FMMVLN model

n	k^{a}	$\widehat{\pi}$ (sd)	$\widehat{M}$ (sd)	$\widehat{\Omega}$ (sd)	$\widehat{\Sigma}$ (sd)
500	1	0.330	$\begin{bmatrix}0.000 & 0.003 & 0.005\\ 0.006 & 0.006 & -0.002\\ 0.008 & -0.001 & 0.003\end{bmatrix}$	$\begin{bmatrix}0.373 & -0.002 & 0.002\\ -0.002 & 0.375 & 0.001\\ 0.002 & 0.001 & 0.373\end{bmatrix}$	$\begin{bmatrix}1.005 & -0.003 & -0.002\\ -0.003 & 0.991 & 0.000\\ -0.002 & 0.000 & 0.992\end{bmatrix}$
		(0.015)	$\begin{pmatrix}\begin{bmatrix}0.048 & 0.047 & 0.049\\ 0.052 & 0.055 & 0.047\\ 0.051 & 0.052 & 0.047\end{bmatrix}\end{pmatrix}$	$\begin{pmatrix}\begin{bmatrix}0.028 & 0.021 & 0.018\\ 0.021 & 0.027 & 0.016\\ 0.018 & 0.016 & 0.023\end{bmatrix}\end{pmatrix}$	$\begin{pmatrix}\begin{bmatrix}0.050 & 0.045 & 0.054\\ 0.045 & 0.056 & 0.052\\ 0.054 & 0.052 & 0.057\end{bmatrix}\end{pmatrix}$
	2	0.332	$\begin{bmatrix}0.704 & 0.700 & 0.707\\ 0.703 & 0.701 & 0.695\\ 0.707 & 0.712 & 0.706\end{bmatrix}$	$\begin{bmatrix}0.379 & -0.001 & 0.003\\ -0.001 & 0.380 & 0.005\\ 0.003 & 0.005 & 0.383\end{bmatrix}$	$\begin{bmatrix}0.996 & 0.000 & 0.003\\ 0.000 & 0.990 & -0.002\\ 0.003 & -0.002 & 0.969\end{bmatrix}$
		(0.018)	$\begin{pmatrix}\begin{bmatrix}0.058 & 0.057 & 0.056\\ 0.061 & 0.061 & 0.052\\ 0.060 & 0.057 & 0.053\end{bmatrix}\end{pmatrix}$	$\begin{pmatrix}\begin{bmatrix}0.028 & 0.019 & 0.019\\ 0.019 & 0.027 & 0.022\\ 0.019 & 0.022 & 0.028\end{bmatrix}\end{pmatrix}$	$\begin{pmatrix}\begin{bmatrix}0.055 & 0.055 & 0.056\\ 0.055 & 0.059 & 0.048\\ 0.056 & 0.048 & 0.062\end{bmatrix}\end{pmatrix}$
	3	0.338	$\begin{bmatrix}1.404 & 1.399 & 1.407\\ 1.402 & 1.395 & 1.406\\ 1.400 & 1.397 & 1.407\end{bmatrix}$	$\begin{bmatrix}0.372 & 0.001 & -0.003\\ 0.001 & 0.374 & 0.001\\ -0.003 & 0.001 & 0.376\end{bmatrix}$	$\begin{bmatrix}0.992 & -0.002 & 0.001\\ -0.002 & 1.001 & 0.001\\ 0.001 & 0.001 & 0.994\end{bmatrix}$
		(0.014)	$\begin{pmatrix}\begin{bmatrix}0.048 & 0.045 & 0.050\\ 0.053 & 0.042 & 0.049\\ 0.058 & 0.050 & 0.049\end{bmatrix}\end{pmatrix}$	$\begin{pmatrix}\begin{bmatrix}0.024 & 0.018 & 0.016\\ 0.018 & 0.028 & 0.019\\ 0.016 & 0.019 & 0.025\end{bmatrix}\end{pmatrix}$	$\begin{pmatrix}\begin{bmatrix}0.049 & 0.044 & 0.050\\ 0.044 & 0.052 & 0.050\\ 0.050 & 0.050 & 0.051\end{bmatrix}\end{pmatrix}$
750	1	0.299	$\begin{bmatrix}0.504 & 0.500 & 0.499\\ 0.502 & 0.503 & 0.507\\ 0.484 & 0.496 & 0.502\end{bmatrix}$	$\begin{bmatrix}0.999 & -0.005 & -0.001\\ -0.005 & 0.993 & -0.003\\ -0.001 & -0.003 & 0.994\end{bmatrix}$	$\begin{bmatrix}0.998 & -0.001 & 0.001\\ -0.001 & 1.000 & 0.001\\ 0.001 & 0.001 & 1.002\end{bmatrix}$
		(0.018)	$\begin{pmatrix}\begin{bmatrix}0.065 & 0.077 & 0.067\\ 0.078 & 0.086 & 0.074\\ 0.090 & 0.078 & 0.081\end{bmatrix}\end{pmatrix}$	$\begin{pmatrix}\begin{bmatrix}0.060 & 0.049 & 0.046\\ 0.049 & 0.063 & 0.041\\ 0.046 & 0.041 & 0.058\end{bmatrix}\end{pmatrix}$	$\begin{pmatrix}\begin{bmatrix}0.048 & 0.041 & 0.045\\ 0.041 & 0.047 & 0.042\\ 0.045 & 0.042 & 0.047\end{bmatrix}\end{pmatrix}$
	2	0.301	$\begin{bmatrix}1.199 & 1.198 & 1.190\\ 1.198 & 1.204 & 1.194\\ 1.192 & 1.203 & 1.203\end{bmatrix}$	$\begin{bmatrix}0.559 & 0.004 & -0.005\\ 0.004 & 0.560 & -0.001\\ -0.005 & -0.001 & 0.558\end{bmatrix}$	$\begin{bmatrix}1.003 & 0.008 & 0.005\\ 0.008 & 0.996 & -0.002\\ 0.005 & -0.002 & 0.998\end{bmatrix}$
		(0.018)	$\begin{pmatrix}\begin{bmatrix}0.067 & 0.058 & 0.060\\ 0.067 & 0.062 & 0.062\\ 0.061 & 0.063 & 0.063\end{bmatrix}\end{pmatrix}$	$\begin{pmatrix}\begin{bmatrix}0.041 & 0.026 & 0.029\\ 0.026 & 0.037 & 0.027\\ 0.029 & 0.027 & 0.040\end{bmatrix}\end{pmatrix}$	$\begin{pmatrix}\begin{bmatrix}0.057 & 0.047 & 0.051\\ 0.047 & 0.057 & 0.047\\ 0.051 & 0.047 & 0.056\end{bmatrix}\end{pmatrix}$
	3	0.400	$\begin{bmatrix}1.993 & 1.998 & 2.002\\ 2.002 & 2.000 & 2.001\\ 1.998 & 2.003 & 1.998\end{bmatrix}$	$\begin{bmatrix}0.253 & 0.001 & 0.001\\ 0.001 & 0.250 & 0.000\\ 0.001 & 0.000 & 0.253\end{bmatrix}$	$\begin{bmatrix}0.993 & 0.002 & -0.002\\ 0.002 & 0.996 & 0.000\\ -0.002 & 0.000 & 0.992\end{bmatrix}$
		(0.005)	$\begin{pmatrix}\begin{bmatrix}0.030 & 0.027 & 0.029\\ 0.033 & 0.030 & 0.030\\ 0.031 & 0.028 & 0.030\end{bmatrix}\end{pmatrix}$	$\begin{pmatrix}\begin{bmatrix}0.012 & 0.009 & 0.008\\ 0.009 & 0.013 & 0.009\\ 0.008 & 0.009 & 0.012\end{bmatrix}\end{pmatrix}$	$\begin{pmatrix}\begin{bmatrix}0.045 & 0.029 & 0.039\\ 0.029 & 0.041 & 0.033\\ 0.039 & 0.033 & 0.042\end{bmatrix}\end{pmatrix}$

[a] represents the component label in mixture model.

The results in Table 3 presents the count of samples (out of 100) where different numbers of components (g) results the best AIC and BIC values. From the results, the highest number of samples achieving optimal AIC and BIC values for the two-component model occurs at $g = 2$, and for the three-component model, at $g = 3$. This indicates that the proposed FMMVLN model effectively identifies the correct number of components according to both AIC and BIC criteria. Further, the extended K-means algorithm with two and three clusters are performed. Using the results of extended K-means as initial values, two and three-component FMMVLN models are fitted and the results are presented in Tables 4 and

5, respectively. The results shows that the estimates are close to the actual parameter values. Here, the covariance matrices can only be estimated up to a multiplicative constant as they depend on each other within the component. The standard deviation of all estimates decrease as the sample sizes increase, confirming the estimates' consistency property.

Table 6: The average MCR and ARI values with standard deviations for two and three-component FMMVLN models

Model	n	MCR (sd)	ARI (sd)
2-component	250	0.100 (0.021)	0.637 (0.067)
	500	0.097 (0.017)	0.650 (0.055)
	750	0.077 (0.010)	0.717 (0.035)
3-component	500	0.065 (0.011)	0.818 (0.028)
	750	0.073 (0.009)	0.819 (0.022)

Table 7: Comparison of convergence speed (number of iterations) for random and extended K-means initialization techniques

Model	n	Random initialization Median (IQR)	K-means initialization Median (IQR)
2-component	250	60.5 (32.50)	43.0 (23.25)
	500	83.0 (38.25)	44.0 (24.25)
	750	46.0 (13.00)	28.0 (8.00)
3-component	500	67.5 (44.50)	22.5 (7.00)
	750	50.0 (37.75)	38.0 (12.00)

The model-based clustering using FMMVLN is performed for each dataset of each sample size, and observations are segregated into clusters corresponding to the maximum posterior probabilities of component densities obtained from the EM algorithm. The MCR and ARI values with their standard deviations are reported in Table 6. From these results, as the sample size increases, the MCR values decrease, and the ARI values increase. The standard deviations of MCR and ARI also decrease as the sample size increases. These results indicate that clustering is more reliable for larger sample sizes.

In order to assess the effectiveness of extended K-means as an initialization method in terms of convergence speed, we also fitted the FMMVLN models using the existing random initialization for each of the simulated datasets, noting the number of iterations required to reach convergence. The average (median) number of iterations with inter-quartile range (IQR) for random initialization and the extended K-means initialization are compared in Table 7. From these results, in all cases, the extended K-means initialization technique outperformed the random initialization by achieving convergence with considerably fewer iterations.

5. Landsat data

To assess the performance of the proposed model, we utilize the Landsat satellite data (Srinivasan, 1993), which is well-known and extensively used in image classification

tasks. This data set is available through the UCI Machine Learning Repository and was originally described by Ashwin Srinivasan in 1993. For this analysis, we focus on a test dataset consisting of 2000 samples. The dataset captures multi-spectral pixel values within a 3×3 neighbourship in satellite images. Each neighbourhood is represented by 36 variables, corresponding to four spectral bands, with each band containing nine pixels. Hence, the dimension of the data is of 4×9. The pixel classes are numerically coded in the range of 0 and 255 with 0 corresponding to black and 255 to white. The data is categorized into seven distinct classes: red soil, cotton crop, grey soil, damp grey soil, soil with vegetation stubble, a mixture class, and very damp grey soil. However, in this particular dataset, only six classes are present, as there are no observations for the mixture class. The objective of the model is to accurately classify these pixels into their respective categories based on the spectral information provided.

Given that the number of classes is predefined in this scenario, the extended K-means algorithm is initially applied, specifying six clusters to correspond with the six classes. The resulting cluster centers from this K-means analysis are then used to initialize the FMMVLN distributions. Following this, a six-component FMMVLN model is fitted to the data. The parameter estimates obtained from both the extended K-means and the FMMVLN distributions are summarized in Tables 10 and 11, respectively. These results reveal that the center matrices produced by the extended K-means closely align with the mean matrices derived from the FMMVLN distributions, indicating that the K-means initialization provides a solid foundation for the subsequent model fitting.

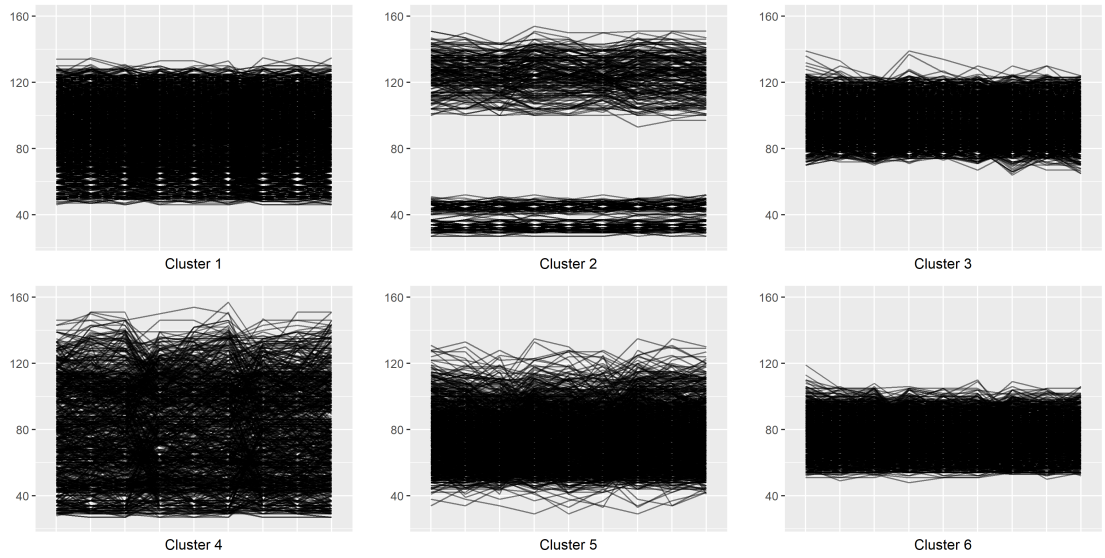

Figure 2: Parallel coordinate plot for the estimated clusters of FMMVLN distributions for Landsat data

However, it is important to note that, as previously discussed, the covariance structures within each component of the mixture model can only be estimated up to a multiplicative constant. As a result, the row covariance matrices Ω_j are estimated to have values close to zero, while the column covariance matrices Σ_j are found to be significantly different from zero. This suggests that while the K-means algorithm effectively initializes the model,

the estimation of covariance matrices requires careful interpretation, particularly given the inherent dependency structures within the model components.

Table 8: Model fit and clustering results of FMMVN, FMMVST, and FMMVLN models

Model	AIC	BIC	ARI
FMMVN	−401435	−404521	0.617
FMMVST	−417662	−421958	0.419
FMMVLN	−400145	−403231	0.638

Table 9: Cluster assignments of (a) FMMVN, (b) FMMVST, and (c) FMMVLN models

(a) FMMVN	Estimated class					
True class	1	2	3	4	5	6
Red soil	442	6	0	12	1	0
Cotton crop	0	222	0	2	0	0
Grey soil	5	3	337	52	0	0
Damp grey soil	1	17	68	29	1	95
Soil with vegetation stubble	9	44	0	62	122	0
Very damp grey soil	0	9	19	106	17	319

(b) FMMVST	Estimated class					
True class	1	2	3	4	5	6
Red soil	433	0	13	14	1	0
Cotton crop	0	197	0	26	1	0
Grey soil	6	0	368	18	5	0
Damp grey soil	1	6	186	16	2	0
Soil with vegetation stubble	22	48	26	63	71	7
Very damp grey soil	2	12	374	17	10	55

(c) FMMVLN	Estimated class					
True class	1	2	3	4	5	6
Red soil	442	0	0	5	14	0
Cotton crop	0	106	0	116	2	0
Grey soil	10	0	354	3	27	3
Damp grey soil	1	0	72	13	15	110
Soil with vegetation stubble	8	0	0	31	197	1
Very damp grey soil	1	0	18	9	78	364

To visually inspect the separability of the estimated clusters, we constructed a parallel coordinate plot presented in Figure 2. In this plot, each observation rows are represented as a line connecting points on parallel axes, with the position of each point corresponding to the value in the columns. We observed substantial overlap between clusters 1, 4, and 5 with clusters 3 and 6, indicating potential similarity or ambiguity in their feature distributions. In contrast, cluster 2 exhibited a distinct pattern with data points occupying two separate value ranges, resulting in a visible gap within the plot.

Additionally, we conducted a comparative analysis by fitting two well-known finite mixture models: the finite mixtures of matrix variate normal (FMMVN) distributions and the finite mixtures of matrix variate Skew-t (FMMVST) distributions. The results of this analysis are presented in Table 8. For this comparison, model-based clustering was performed using the posterior probabilities obtained from each of these finite mixture models, and the ARI values were calculated to evaluate the accuracy of the clustering. The AIC and BIC values were also computed for each model to assess their fit to the data.

The results in Table 8 shows that the proposed model outperformed the FMMVN and FMMVST models, as indicated by its higher AIC and BIC values. This suggests that the proposed model provides a better fit to the data compared to the other two models. Furthermore, the ARI values confirm that the proposed model is more effective in correctly classifying a greater number of pixels into their respective categories, demonstrating its superior performance in this clustering task. Despite the low increment in the ARI value, likely due to substantial overlap between clusters as visualized in the parallel coordinate plot, the proposed model demonstrates superior performance in fitting and clustering skewed matrix variate data overall. We obtained the cluster assignment tables for the three finite mixtures and presented in Table 9. Since the clusters don't follow a specific sequence, the cluster assignment table can only show which observations are assigned to each cluster. However, these tables offer information on how the observations in estimated clusters are distributed among the true classes.

6. Results and discussion

In this paper, we made an attempt to model and cluster the skewed three-way data with underlying sub-populations using FMMVLN distributions. To initialize the FMMVLN models, we introduced an extended K-means algorithm. The subsequent parameter estimation was achieved through the EM algorithm. We employed AIC and BIC to determine the optimal number of components and the model fit in the mixture models. The results demonstrated that BIC exhibited superior performance in accurately identifying the correct number of components compared to AIC. Furthermore, the clustering performance, evaluated using the MCR and ARI values. The computations of simulation and real data studies are carried out in R.

The simulation study findings revealed that the FMMVLN models effectively estimated the true parameter values, with estimation accuracy improving as sample size increased. This confirms the consistency property of the estimators. Furthermore, the clustering performance, evaluated using MCR and ARI demonstrated significant improvements with increasing sample size. This indicates that the proposed FMMVLN-based clustering approach is more reliable for larger datasets. In addition, a comparative analysis of initial-

ization methods highlighted the efficiency of the extended K-means algorithm in terms of convergence speed. The proposed initialization consistently outperformed random initialization, leading to a substantial reduction in the number of iterations required for the EM algorithm to converge.

The real data study reveals that the resulting parameter estimates from extended K-means and FMMVLN model exhibited strong alignment, indicating the effectiveness of the K-means initialization. The comparative analysis with FMMVN and FMMVST models demonstrated the superior performance of the proposed FMMVLN model in terms of model fit and clustering accuracy, as evidenced by higher AIC, BIC, and ARI values. Despite some overlap between clusters, the FMMVLN model effectively handled the skewed matrix variate data, outperforming the alternative models.

One major challenge in finite mixture models for matrix variate data is the initialization of the EM algorithm. To address this, we proposed the extended K-means algorithm to cluster the matrix variate data. Additionally, estimating covariance matrices is complex due to inherent dependencies within the model components. Furthermore, these models often require large sample sizes as the number of parameters increases with data dimensionality and the number of mixture components. We aim to explore these issues further in future research.

Acknowledgements

We thank the editors and the reviewers for their valuable comments and suggestions, which helped improve this article's work.

Conflict of interest

The authors do not have any financial or non-financial conflict of interest to declare for the research work included in this article.

References

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, **19**, 716–723.

Biernacki, C., Celeux, G., and Govaert, G. (2003). Choosing starting values for the em algorithm for getting the highest likelihood in multivariate Gaussian mixture models. *Computational Statistics & Data Analysis*, **41**, 561–575.

Bouguila, N. and ElGuebaly, W. (2009). Discrete data clustering using finite mixture models. *Pattern Recognition*, **42**, 33–42.

Deepana, R. and Kiruthika (2018). Model based clustering using finite mixture models of lognormal distribution. *Research & Reviews: Journal of Statistics*, **7**, 58–67.

Deepana, R. and Kiruthika, C. (2022). Clustering using skewed data via finite mixtures of multivariate lognormal distributions. *Statistics and Applications*, **20**, 219–237.

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)*, **39**, 1–38.

Doğru, F. Z., Bulut, Y. M., and Arslan, O. (2016). Finite mixtures of matrix variate t distributions. *Gazi University Journal of Science*, **29**, 335–341.

Everitt, B. (2013). *Finite Mixture Distributions*. Springer Science & Business Media.

Gallaugher, M. P. and McNicholas, P. D. (2018). Finite mixtures of skewed matrix variate distributions. *Pattern Recognition*, **80**, 83–93.

Hu, Z. (2015). *Initializing the EM Algorithm for Data Clustering and Sub-population Detection*. PhD thesis, The Ohio State University.

Karlis, D. and Xekalaki, E. (2003). Choosing initial values for the em algorithm for finite mixtures. *Computational Statistics & Data Analysis*, **41**, 577–590.

McLachlan, G. J., Lee, S. X., and Rathnayake, S. I. (2019). Finite mixture models. *Annual Review of Statistics and Its Application*, **6**, 355–378.

McNicholas, P. D. and Murphy, T. B. (2008). Parsimonious Gaussian mixture models. *Statistics and Computing*, **18**, 285–296.

Melnykov, V. and Maitra, R. (2010). Finite mixture models and model-based clustering. *Statistics Surveys*, **4**, 80–116.

Michael, S. and Melnykov, V. (2016). An effective strategy for initializing the em algorithm in finite mixture models. *Advances in Data Analysis and Classification*, **10**, 563–583.

Panić, B., Klemenc, J., and Nagode, M. (2020). Improved initialization of the em algorithm for mixture model parameter estimation. *Mathematics*, **8**, 373.

Peel, D. and MacLahlan, G. (2000). *Finite Mixture Models*. John Wiley & Sons.

Rand, W. M. (1971). Objective criteria for the evaluation of clustering methods. *Journal of the American Statistical Association*, **66**, 846–850.

Rezaei, A. and Ahmadi, K. (2023). Generalized Mahalanobis distance and its application in detecting matrix outliers. *Filomat*, **37**, 7993–8011.

Rohde, D., Corcoran, J., White, G., and Huang, R. (2012). Visualization of predictive distributions for discrete spatial-temporal log cox processes approximated with mcmc. In *Intelligent Data Engineering and Automated Learning-IDEAL 2012: 13th International Conference, Natal, Brazil, August 29-31, 2012. Proceedings 13*, pages 286–293. Springer.

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, **6**, 461–464.

Silva, A., Qin, X., Rothstein, S. J., McNicholas, P. D., and Subedi, S. (2023). Finite mixtures of matrix variate poisson-log normal distributions for three-way count data. *Bioinformatics*, **39**, 1–8.

Srinivasan, A. (1993). Statlog (landsat satellite). UCI Machine Learning Repository.

Tomarchio, S. D., Gallaugher, M. P., Punzo, A., and McNicholas, P. D. (2022). Mixtures of matrix-variate contaminated normal distributions. *Journal of Computational and Graphical Statistics*, **31**, 413–421.

Tomarchio, S. D., Punzo, A., and Bagnato, L. (2020). Two new matrix-variate distributions with application in model-based clustering. *Computational Statistics & Data Analysis*, **152**, 107050.

Viroli, C. (2011a). Finite mixtures of matrix normal distributions for classifying three-way data. *Statistics and Computing*, **21**, 511–522.

Viroli, C. (2011b). Model based clustering for three-way data structures. *Bayesian Anal.*, **6**, 573–602.

ANNEXURE

Table 10: Estimates of cluster centres and cluster sizes of extended K-means for Landsat data

C [a]	$\widehat{\lambda}$ [b]	$\widehat{c}$
1	0.17	$\begin{bmatrix} 4.13 & 4.13 & 4.13 & 4.13 & 4.12 & 4.12 & 4.13 & 4.12 & 4.12 \\ 4.18 & 4.18 & 4.19 & 4.18 & 4.17 & 4.18 & 4.18 & 4.17 & 4.17 \\ 4.33 & 4.32 & 4.33 & 4.32 & 4.30 & 4.32 & 4.33 & 4.32 & 4.32 \\ 4.09 & 4.08 & 4.09 & 4.08 & 4.07 & 4.08 & 4.10 & 4.09 & 4.10 \end{bmatrix}$
2	0.10	$\begin{bmatrix} 3.83 & 3.83 & 3.84 & 3.83 & 3.82 & 3.82 & 3.85 & 3.84 & 3.84 \\ 3.54 & 3.54 & 3.57 & 3.53 & 3.51 & 3.53 & 3.57 & 3.56 & 3.56 \\ 4.75 & 4.76 & 4.75 & 4.76 & 4.76 & 4.76 & 4.74 & 4.74 & 4.74 \\ 4.81 & 4.81 & 4.79 & 4.82 & 4.83 & 4.82 & 4.80 & 4.80 & 4.79 \end{bmatrix}$
3	0.12	$\begin{bmatrix} 4.05 & 4.05 & 4.05 & 4.05 & 4.05 & 4.05 & 4.05 & 4.05 & 4.05 \\ 4.29 & 4.28 & 4.28 & 4.28 & 4.28 & 4.27 & 4.28 & 4.27 & 4.27 \\ 4.56 & 4.55 & 4.55 & 4.55 & 4.55 & 4.56 & 4.55 & 4.56 & 4.56 \\ 4.40 & 4.40 & 4.41 & 4.39 & 4.40 & 4.41 & 4.40 & 4.41 & 4.41 \end{bmatrix}$
4	0.21	$\begin{bmatrix} 4.29 & 4.28 & 4.27 & 4.28 & 4.28 & 4.27 & 4.28 & 4.28 & 4.27 \\ 4.43 & 4.41 & 4.40 & 4.42 & 4.42 & 4.41 & 4.42 & 4.41 & 4.41 \\ 4.49 & 4.48 & 4.48 & 4.49 & 4.48 & 4.47 & 4.48 & 4.48 & 4.47 \\ 4.25 & 4.25 & 4.25 & 4.25 & 4.24 & 4.24 & 4.25 & 4.24 & 4.24 \end{bmatrix}$
5	0.16	$\begin{bmatrix} 4.21 & 4.21 & 4.20 & 4.21 & 4.20 & 4.19 & 4.19 & 4.19 & 4.19 \\ 4.62 & 4.63 & 4.61 & 4.62 & 4.63 & 4.62 & 4.61 & 4.62 & 4.61 \\ 4.73 & 4.74 & 4.73 & 4.74 & 4.74 & 4.73 & 4.73 & 4.74 & 4.73 \\ 4.53 & 4.53 & 4.52 & 4.53 & 4.53 & 4.53 & 4.53 & 4.53 & 4.52 \end{bmatrix}$
6	0.24	$\begin{bmatrix} 4.45 & 4.46 & 4.45 & 4.46 & 4.46 & 4.46 & 4.45 & 4.46 & 4.45 \\ 4.64 & 4.64 & 4.63 & 4.64 & 4.65 & 4.64 & 4.63 & 4.64 & 4.63 \\ 4.69 & 4.69 & 4.68 & 4.69 & 4.69 & 4.69 & 4.68 & 4.69 & 4.68 \\ 4.45 & 4.45 & 4.45 & 4.46 & 4.46 & 4.45 & 4.45 & 4.46 & 4.45 \end{bmatrix}$

[a] represents the cluster label and [b] represents the proportion of cluster size.

Table 11: Parameter estimates of FMMVLN distributions for Landsat data

k^{a}	$\widehat{\pi}$	$\widehat{M}$ / $\widehat{\Omega}$	$\widehat{\Sigma}$
1	0.17	$\begin{bmatrix} 4.18 & 4.16 & 4.15 & 4.18 & 4.16 & 4.16 & 4.19 & 4.18 & 4.16 \\ 4.25 & 4.24 & 4.22 & 4.26 & 4.24 & 4.23 & 4.27 & 4.25 & 4.24 \\ 4.45 & 4.43 & 4.42 & 4.45 & 4.43 & 4.43 & 4.46 & 4.45 & 4.44 \\ 4.24 & 4.23 & 4.22 & 4.24 & 4.23 & 4.22 & 4.25 & 4.24 & 4.24 \end{bmatrix}$ / $\begin{bmatrix} 0.01 & 0.02 & 0.01 & 0.01 \\ 0.02 & 0.03 & 0.01 & 0.01 \\ 0.01 & 0.01 & 0.02 & 0.02 \\ 0.01 & 0.01 & 0.02 & 0.02 \end{bmatrix}$	$\begin{bmatrix} 1.09 & 0.87 & 0.63 & 0.82 & 0.75 & 0.58 & 0.59 & 0.60 & 0.51 \\ 0.87 & 1.11 & 0.84 & 0.73 & 0.81 & 0.73 & 0.53 & 0.60 & 0.58 \\ 0.63 & 0.84 & 1.05 & 0.59 & 0.71 & 0.78 & 0.45 & 0.53 & 0.57 \\ 0.82 & 0.73 & 0.59 & 1.07 & 0.87 & 0.66 & 0.83 & 0.78 & 0.64 \\ 0.75 & 0.81 & 0.71 & 0.87 & 1.11 & 0.87 & 0.76 & 0.85 & 0.79 \\ 0.58 & 0.73 & 0.78 & 0.66 & 0.87 & 1.08 & 0.63 & 0.77 & 0.84 \\ 0.59 & 0.53 & 0.45 & 0.83 & 0.76 & 0.63 & 1.11 & 0.89 & 0.70 \\ 0.60 & 0.60 & 0.53 & 0.78 & 0.85 & 0.77 & 0.89 & 1.13 & 0.90 \\ 0.51 & 0.58 & 0.57 & 0.64 & 0.79 & 0.84 & 0.70 & 0.90 & 1.13 \end{bmatrix}$
2	0.05	$\begin{bmatrix} 3.80 & 3.80 & 3.81 & 3.80 & 3.81 & 3.81 & 3.81 & 3.81 & 3.82 \\ 3.48 & 3.48 & 3.49 & 3.47 & 3.48 & 3.48 & 3.49 & 3.49 & 3.50 \\ 4.80 & 4.79 & 4.78 & 4.80 & 4.79 & 4.79 & 4.78 & 4.78 & 4.78 \\ 4.87 & 4.86 & 4.85 & 4.86 & 4.86 & 4.85 & 4.86 & 4.85 & 4.84 \end{bmatrix}$ / $\begin{bmatrix} 0.00 & 0.00 & 0.00 & 0.00 \\ 0.00 & 0.01 & 0.00 & 0.00 \\ 0.00 & 0.00 & 0.00 & 0.00 \\ 0.00 & 0.00 & 0.00 & 0.00 \end{bmatrix}$	$\begin{bmatrix} 1.24 & 0.90 & 0.79 & 0.90 & 0.82 & 0.72 & 0.76 & 0.76 & 0.67 \\ 0.90 & 1.20 & 0.95 & 0.87 & 0.78 & 0.76 & 0.69 & 0.68 & 0.65 \\ 0.79 & 0.95 & 1.25 & 0.80 & 0.77 & 0.81 & 0.65 & 0.66 & 0.64 \\ 0.90 & 0.87 & 0.80 & 1.41 & 1.03 & 0.85 & 0.98 & 1.00 & 0.90 \\ 0.82 & 0.78 & 0.77 & 1.03 & 1.23 & 0.89 & 0.87 & 0.87 & 0.81 \\ 0.72 & 0.76 & 0.81 & 0.85 & 0.89 & 1.12 & 0.80 & 0.79 & 0.76 \\ 0.76 & 0.69 & 0.65 & 0.98 & 0.87 & 0.80 & 1.45 & 1.13 & 0.94 \\ 0.76 & 0.68 & 0.66 & 1.00 & 0.87 & 0.79 & 1.13 & 1.41 & 1.08 \\ 0.67 & 0.65 & 0.64 & 0.90 & 0.81 & 0.76 & 0.94 & 1.08 & 1.35 \end{bmatrix}$
3	0.09	$\begin{bmatrix} 4.01 & 4.01 & 4.02 & 4.00 & 4.00 & 4.00 & 4.00 & 4.00 & 4.00 \\ 3.93 & 3.93 & 3.94 & 3.93 & 3.91 & 3.91 & 3.92 & 3.91 & 3.92 \\ 4.62 & 4.63 & 4.63 & 4.62 & 4.63 & 4.64 & 4.63 & 4.63 & 4.63 \\ 4.56 & 4.57 & 4.57 & 4.57 & 4.58 & 4.59 & 4.58 & 4.58 & 4.58 \end{bmatrix}$ / $\begin{bmatrix} 0.03 & 0.05 & -0.01 & -0.03 \\ 0.05 & 0.11 & -0.02 & -0.06 \\ -0.01 & -0.02 & 0.02 & 0.02 \\ -0.03 & -0.06 & 0.02 & 0.05 \end{bmatrix}$	$\begin{bmatrix} 0.95 & 0.79 & 0.61 & 0.65 & 0.65 & 0.51 & 0.36 & 0.40 & 0.41 \\ 0.79 & 0.98 & 0.86 & 0.56 & 0.66 & 0.67 & 0.29 & 0.34 & 0.43 \\ 0.61 & 0.86 & 1.10 & 0.49 & 0.60 & 0.71 & 0.23 & 0.30 & 0.41 \\ 0.65 & 0.56 & 0.49 & 0.85 & 0.68 & 0.50 & 0.57 & 0.55 & 0.46 \\ 0.65 & 0.66 & 0.60 & 0.68 & 0.91 & 0.72 & 0.47 & 0.59 & 0.64 \\ 0.51 & 0.67 & 0.71 & 0.50 & 0.72 & 0.94 & 0.40 & 0.51 & 0.67 \\ 0.36 & 0.29 & 0.23 & 0.57 & 0.47 & 0.40 & 0.86 & 0.66 & 0.49 \\ 0.40 & 0.34 & 0.30 & 0.55 & 0.59 & 0.51 & 0.66 & 0.87 & 0.72 \\ 0.41 & 0.43 & 0.41 & 0.46 & 0.64 & 0.67 & 0.49 & 0.72 & 0.97 \end{bmatrix}$

Table 11: Continued

k^{a}	$\widehat{\pi}$	$\widehat{M}$	$\widehat{\Omega}$	$\widehat{\Sigma}$
4	0.24	$\begin{bmatrix} 4.25 & 4.25 & 4.25 & 4.25 & 4.24 & 4.24 & 4.24 & 4.24 & 4.24 \\ 4.38 & 4.37 & 4.38 & 4.37 & 4.37 & 4.37 & 4.37 & 4.37 & 4.37 \\ 4.42 & 4.42 & 4.42 & 4.42 & 4.41 & 4.42 & 4.42 & 4.41 & 4.41 \\ 4.17 & 4.17 & 4.17 & 4.17 & 4.17 & 4.17 & 4.17 & 4.17 & 4.17 \end{bmatrix}$	$\begin{bmatrix} 0.01 & 0.00 & 0.00 & 0.00 \\ 0.00 & 0.01 & 0.01 & 0.01 \\ 0.00 & 0.01 & 0.01 & 0.01 \\ 0.00 & 0.01 & 0.01 & 0.01 \end{bmatrix}$	$\begin{bmatrix} 0.77 & 0.50 & 0.39 & 0.45 & 0.42 & 0.37 & 0.35 & 0.34 & 0.33 \\ 0.50 & 0.75 & 0.50 & 0.40 & 0.43 & 0.42 & 0.34 & 0.34 & 0.35 \\ 0.39 & 0.50 & 0.77 & 0.33 & 0.36 & 0.41 & 0.28 & 0.29 & 0.33 \\ 0.45 & 0.40 & 0.33 & 0.75 & 0.48 & 0.36 & 0.44 & 0.40 & 0.34 \\ 0.42 & 0.43 & 0.36 & 0.48 & 0.70 & 0.45 & 0.39 & 0.41 & 0.39 \\ 0.37 & 0.42 & 0.41 & 0.36 & 0.45 & 0.69 & 0.34 & 0.36 & 0.40 \\ 0.35 & 0.34 & 0.28 & 0.44 & 0.39 & 0.34 & 0.74 & 0.47 & 0.38 \\ 0.34 & 0.34 & 0.29 & 0.40 & 0.41 & 0.36 & 0.47 & 0.69 & 0.45 \\ 0.33 & 0.35 & 0.33 & 0.34 & 0.39 & 0.40 & 0.38 & 0.45 & 0.69 \end{bmatrix}$
5	0.23	$\begin{bmatrix} 4.16 & 4.15 & 4.15 & 4.15 & 4.15 & 4.14 & 4.14 & 4.14 & 4.14 \\ 4.55 & 4.55 & 4.55 & 4.55 & 4.55 & 4.55 & 4.55 & 4.55 & 4.54 \\ 4.68 & 4.68 & 4.68 & 4.68 & 4.68 & 4.68 & 4.67 & 4.68 & 4.68 \\ 4.48 & 4.48 & 4.48 & 4.48 & 4.48 & 4.48 & 4.48 & 4.48 & 4.48 \end{bmatrix}$	$\begin{bmatrix} 0.01 & 0.01 & 0.00 & 0.00 \\ 0.01 & 0.01 & 0.01 & 0.01 \\ 0.00 & 0.01 & 0.01 & 0.00 \\ 0.00 & 0.01 & 0.00 & 0.01 \end{bmatrix}$	$\begin{bmatrix} 1.18 & 0.93 & 0.73 & 0.86 & 0.78 & 0.64 & 0.71 & 0.65 & 0.56 \\ 0.93 & 1.17 & 0.91 & 0.81 & 0.82 & 0.77 & 0.67 & 0.66 & 0.63 \\ 0.73 & 0.91 & 1.15 & 0.70 & 0.78 & 0.84 & 0.59 & 0.64 & 0.68 \\ 0.86 & 0.81 & 0.70 & 1.13 & 0.87 & 0.69 & 0.83 & 0.75 & 0.64 \\ 0.78 & 0.82 & 0.78 & 0.87 & 1.06 & 0.84 & 0.76 & 0.78 & 0.74 \\ 0.64 & 0.77 & 0.84 & 0.69 & 0.84 & 1.10 & 0.67 & 0.74 & 0.80 \\ 0.71 & 0.67 & 0.59 & 0.83 & 0.76 & 0.67 & 1.10 & 0.84 & 0.70 \\ 0.65 & 0.66 & 0.64 & 0.75 & 0.78 & 0.74 & 0.84 & 1.03 & 0.82 \\ 0.56 & 0.63 & 0.68 & 0.64 & 0.74 & 0.80 & 0.70 & 0.82 & 1.06 \end{bmatrix}$
6	0.22	$\begin{bmatrix} 4.46 & 4.46 & 4.46 & 4.46 & 4.46 & 4.46 & 4.46 & 4.46 & 4.46 \\ 4.64 & 4.64 & 4.64 & 4.65 & 4.64 & 4.64 & 4.64 & 4.64 & 4.64 \\ 4.69 & 4.69 & 4.68 & 4.69 & 4.69 & 4.69 & 4.68 & 4.69 & 4.68 \\ 4.45 & 4.45 & 4.45 & 4.45 & 4.45 & 4.45 & 4.45 & 4.45 & 4.45 \end{bmatrix}$	$\begin{bmatrix} 0.00 & 0.00 & 0.00 & 0.00 \\ 0.00 & 0.00 & 0.00 & 0.00 \\ 0.00 & 0.00 & 0.00 & 0.00 \\ 0.00 & 0.00 & 0.00 & 0.00 \end{bmatrix}$	$\begin{bmatrix} 0.69 & 0.39 & 0.30 & 0.36 & 0.29 & 0.24 & 0.27 & 0.22 & 0.18 \\ 0.39 & 0.64 & 0.40 & 0.33 & 0.32 & 0.29 & 0.28 & 0.24 & 0.22 \\ 0.30 & 0.40 & 0.69 & 0.31 & 0.33 & 0.34 & 0.25 & 0.26 & 0.29 \\ 0.36 & 0.33 & 0.31 & 0.66 & 0.39 & 0.31 & 0.37 & 0.32 & 0.26 \\ 0.29 & 0.32 & 0.33 & 0.39 & 0.64 & 0.40 & 0.33 & 0.33 & 0.30 \\ 0.24 & 0.29 & 0.34 & 0.31 & 0.40 & 0.68 & 0.31 & 0.35 & 0.38 \\ 0.27 & 0.28 & 0.25 & 0.37 & 0.33 & 0.31 & 0.72 & 0.40 & 0.30 \\ 0.22 & 0.24 & 0.26 & 0.32 & 0.33 & 0.35 & 0.40 & 0.65 & 0.42 \\ 0.18 & 0.22 & 0.29 & 0.26 & 0.30 & 0.38 & 0.30 & 0.42 & 0.76 \end{bmatrix}$

[a] represents the component label in mixture model.